DETROIT
IEEE/RSJ International Conference on Intelligent Robots and Systems
OCTOBER 1-5, 2023

Real-time Simultaneous Multi-Object 3D Shape Reconstruction, 6DoF Pose Estimation and Dense Grasp Prediction

Shubham Agrawal, Nikhil Chavan-Dafle, Isaac Kasahara, Selim Engin, Jinwook Huh, Volkan Isler

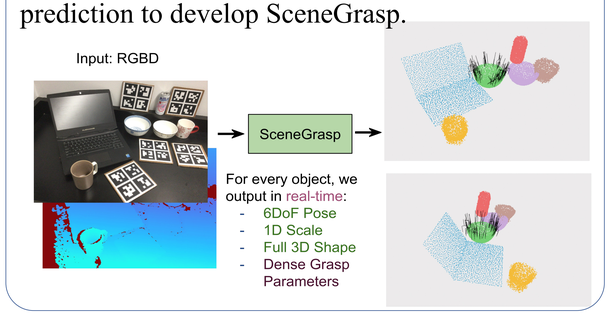
Samsung AI Center New York

Motivation

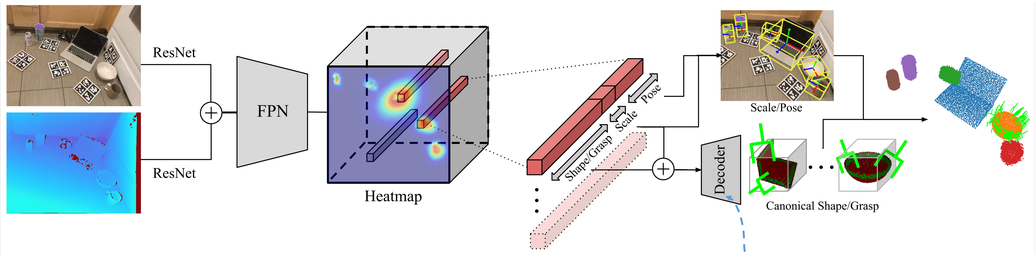
Semantic scene understanding guides the robot action planning; for example:

- Task planning
- Grasp prediction
- Collision-free motion planning

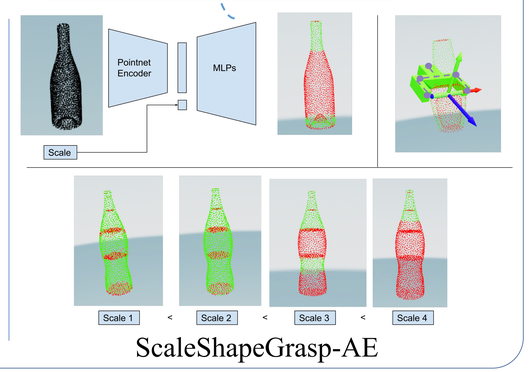
We study the interdependence of scene and action prediction to develop SceneGrasp.



Method Overview



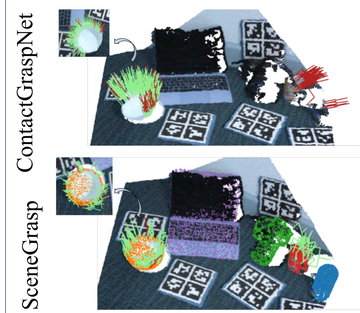
- Dense object descriptors are sampled from the predicted heatmap
- Object descriptor has (a) camera frame object pose, (b) scale, and (c) shape-grasp embedding
- Shape-grasp embedding along with scale is passed through ScaleShapeGrasp-AE to obtain canonical 3D shape and camera frame grasps parameters



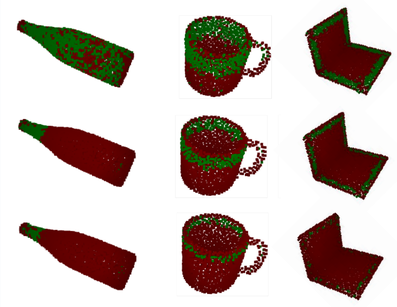
ScaleShapeGrasp-AE

Grasp Prediction Results

Comparison of grasp predictions



Grasp prediction dependency on scale

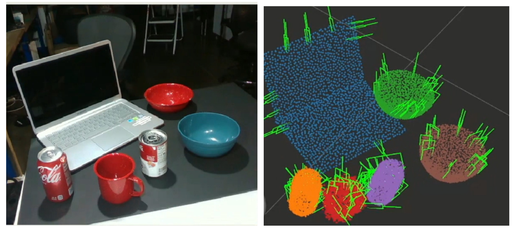


Using estimated shape reconstruction and object scale, SceneGrasp generates more accurate grasp proposals and avoids false positives.

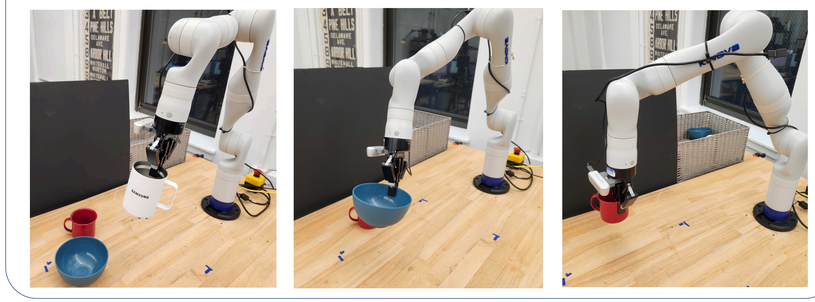
Real-world Experiments

- SceneGrasp enables scene and grasp prediction at **30FPS**
- SceneGrasp works with cluttered scenes involving multiple partially occluded objects.

Shape and grasp predictions in a cluttered scene



Robot execution of a table clean-up task using SceneGrasp



Shape Reconstruction Results

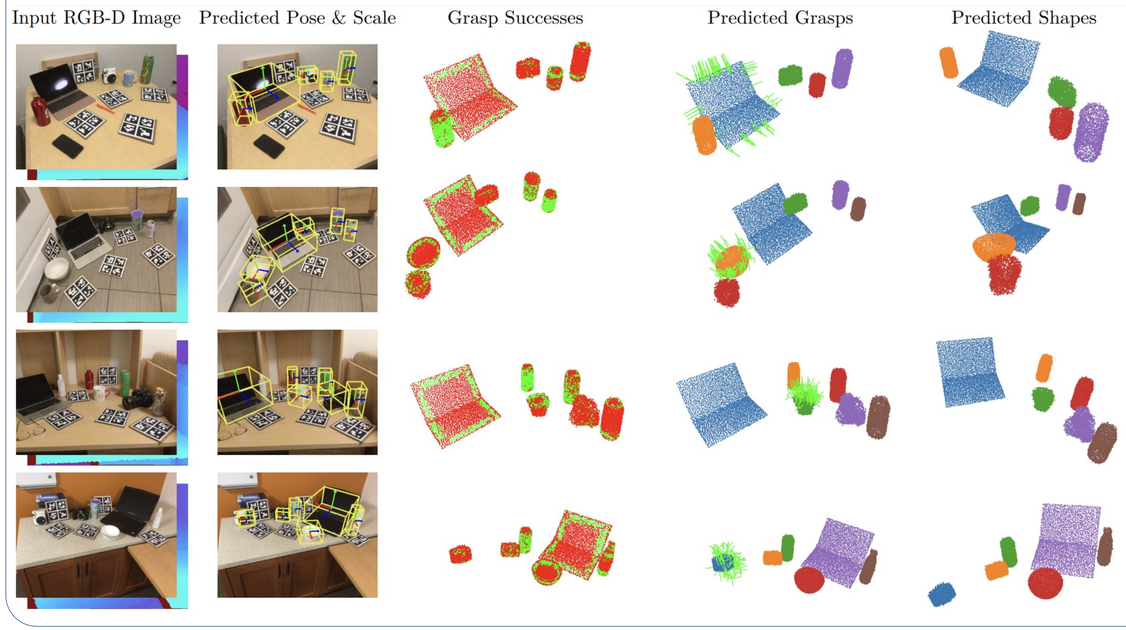


Table 1: **NOCS 3D shape reconstruction** evaluated with Chamfer distance metric (10^{-2} m)

Object	DeformNet	CenterSnap	Ours
Bottle	0.50	**0.13**	0.19
Bowl	0.12	0.10	**0.09**
Camera	0.99	0.43	**0.41**
Can	0.24	**0.09**	**0.09**
Laptop	0.71	**0.07**	0.13
Mug	0.10	**0.06**	0.17

Table 2: **NOCS 3D pose estimation** (IOU precision)

Method	IOU25	IOU50	5°5cm	5°10cm	10°5cm	10°10cm
NOCS	**84.8**	78.0	10.0	9.8	25.2	25.8
CenterSnap	83.5	**80.2**	**27.2**	**29.2**	**58.8**	**64.4**
Ours	79.7	76.2	18.4	22.5	37.5	54.9

SceneGrasp shows comparable performance to the baselines, dedicated for shape reconstruction and pose estimation, while simulaneously generating grasp proposals.

Key Takeaway

SceneGrasp is a method for **simultaneous** object-level scene understanding, i.e., **object classification, reconstruction, and pose estimation**, and **dense grasp estimation** of multiple objects from a single view RGBD image which runs in **real-time at 30FPS**.